Int. j. inf. tecnol.
https://doi.org/10.1007/s41870-020-00569-8

ORIGINAL RESEARCH

CMBA-SVM: a clinical approach for Parkinson disease diagnosis

Bibhuprasad Sahu[1] · **Sachi Nandan Mohanty**[2]

Received: 18 July 2020 / Accepted: 9 November 2020

Abstract Different intelligence models are used by researchers for an easy and successful diagnosis of neurodegenerative diseases like Parkinson's disease (PD) but none of the adopted methods is efficient. Early-stage identification of disease and diagnosis based upon vocal measurements is important to enhance the productive lives of the patient. An innovative intelligence model with a combination of a chaos-mapped bat algorithm (CMBA) and a support vector machine (SVM) called CMBA-SVM is introduced. The coordination of the CMBA method effectively resolved the SVM parameter tuning issues. CMBA is used for the identification of featured and consider input for SVM to develop an intelligence model. The effectiveness and performance of the proposed model are compared to ECFA-SVM and CFA-SVM. The performance of the proposed model was tested with considering various parameters in terms of accuracy, sensitivity, specificity, and AUC by taking two sets of PD data set (from UCI repository and Istanbul collected data set). The performance results and perspective of this study were compared with the different intelligence methods described literature study that used the same data, and the CMBA-SVM showed better efficiency than the other. The CMBA-SVM has a very good prospect, which may provide huge expediency to the clinicians to improve quality during PD diagnosis.

✉ Bibhuprasad Sahu
prasadnikhil176@gmail.com

[1] Department of Computer Science and Engineering, Gandhi Institute For Technology, Bhubaneswar, Odisha, India

[2] Department of Computer Science and Engineering, IcfaiTech, ICFAI Foundation for higher Education, Hyderabad, Telangana, India

Keywords PD · CMBA · SVM · Chaotic Map · ROC

1 Introduction

PD is a nervous system-related degenerative disease found in human beings due to brain cell loss which impacts the mobility of the body. After Alzheimer's disease, this disease is treated as the second-highest affected by human beings due to the central nervous system disorder. Origin of disorder starts due to the drop-down of the nervous generated chemical known as dopamine, so the message generated for the brain unable to send to the particular part of the body. PD disease is one of the hereditary based on nature. The ratio of the affected patient in the said disease increases rapidly throughout the world, especially in Asian countries. The vocal disorder is the main symptom of detected five years before the diagnosis of PD. The Impairment symptoms belong to PD can be classified into two types such as dysphonia and dysarthria. In the early stage diagnosis identification of dysphonic indicators plays a major role to live quality lives for as long as possible. Later and end-stage identification of the disease may disturb the daily routine of the patient and seeks help from others to live a smooth life. In the advanced stage of the disease various symptom such as legs stiffness, impossible to stand, or walk. Curse of dimensionality is the main concern while designing an intelligence model for the medical data sets. To overcome the course of dimensionality issues we have to identify an efficient feature selection method for the detection of featured gene sets from the huge data sets. Two basic categories of feature selection approaches are used such as filter and wrapper. As the filter approaches do not use learning algorithms, it is very simple and less computational cost. Whereas Wrapper methods

Published online: 11 January 2021

used the learning mechanism to train the model for the feature evaluation and provide better solution comparison with Filter but computational cost is high. In medical data analysis selection criteria of the feature is important as small changes in learning parameter significant variation in the efficiency of the problem. Normally most of the researchers prefer the wrapper approach with the inclusion of meta-heuristic algorithms to handle high dimensional data sets like medical data. Optimal feature subsets are identified using the wrapper method can enhance the efficiency of the classification. Some researchers also implemented the various nature-inspired algorithms to detect the featured subset genes and with the less computational cost. Two basic types of meta-heuristic algorithms are used such as trajectory-based and population-based and population-based meta-heuristic algorithms classified into two types such as evolutionary algorithm and swarm intelligence. Trajectory based meta-heuristic algorithms function will single candidate solution and modification is one in each step until it found the best solution from the neighbor. In this approach to solve a given problem, the search is characterized by a trajectory in the space of candidate solutions. As compare to population-based it is faster and efficient as it handles a single solution at a time. They simply used the exploitation-oriented methods to find out local optimum and promoting the intensification in the search space. Here the algorithm performance constraint on dependent parameters and iteration no's and population-based meta-heuristic wok with multiple candidate solutions. Here the new solution is generated using the recombination of some main features of the original solution. Overall the best solution identified using quality base criterion and exploration methods [1–7].

Different techniques are adopted by various researchers such as ANNs, SVM, Dirichlet process mixtures, similarity classifier, rotation forest, fuzzy k-nearest neighbor (FKNN) to enhance the diagnosis accuracy. SVM has revealed an extremely encouraging tool for diagnosing PD. Bat algorithm (BA) was proposed by Xin-She Yang and its inspiration based on the concept of the echolocation of microbats. Various researchers implemented BA different applications for the optimization of a given problem such as structural optimization, training sessions for sports, detection of featured genes form huge data set, ANNs, operating cost optimization, etc. The random initialization population of BA may lack diversity in the search space. For a small and simple problem, inertia weights are used to control local search step size which may improve the algorithm searching efficiency. But this fails while considering the complex problem, increment of step sizes are indirectly proportional to the diversity agents. Here the concept of inertia weight is introduced and chaotic mapping is used to improve the bat population quality to jump out from the local optimum.Neuro-imaging technique is used by [8] for analyze the performance of PD patients. Chaotic mapping is used to improve the searching mechanism to reach up to the global optimum. We have considered the logistic chaos model of the bat and the mapping number is controlled with the threshold (m) to improve the bat population in each step. But the prone part of the BAT algorithm is the searching phase is limited and may lead towards the local optimum. And a single chaotic map unable to maintain the difference between diversification and intensification phases [9–12]. A chaotic based bat algorithm derived from the inertia weight was proposed. Swarm quality improvement can be done by setting the threshold (m) to control the chaotic mapping time. So the important part is whenever the population falls into the local optimum we can combine the infinite no of the chaotic map to improve bat population. The identification of an efficient learning algorithm is important for the detection of significant gene subsets. From the entire population space, the fitness value of all candidate solutions is evaluated using RBF kernel-SVM learning method. The remainder of the research is arranged as mentioned below: Next section presents a brief description of the background of the logistic map based CBA, RBF kernel-based SVM. And a detailed description of the proposed model CMBA and detailed study about RBF kernel-based SVM presented in Sects. 3 and 4 respectively. Proposed model and experimental results of CMBA-SVM on PD data sets is done in Sects. 5 and 6. In the end, we concluded the study in Sect. 7 with future scope [13, 14].

2 Chaotic maps

Chaos is applied to various fields such as physics, computer science, Bioinformatics, physics, economics, biology, and philosophy. For heuristic optimization solution, numerical problem analysis, decision making analysis, and complex simulating solution, generation of random sequence generation with good consistency is a mandatory consideration. Reduction of storage and computational time leads to enhance to reach up to the need accuracy. Chaos map is a deterministic, random, and nonlinear dynamic system which is non-bound and non-converging [15–17]. Naturally, the efficiency of the Chaos map depends on the initial condition and parameter. It is applied with various optimization problems to improve global convergence without reaching local-minima. As compared to the standard stochastic search, the chaos algorithm carries out the iterative search step very fast due to ergodicity, semi-randomness, and mixing property. To achieve improvement in accuracy with global optimality, we have considered a new chaotic based bat algorithm with SVM. In comparison with

various chaotic mapping models, for mapping the population we have adopted a logistic chaos model. Logistic chaos mapping can be expressed as

$$\beta_{i+1} = \mu\beta_i \times (1 - \beta_i), i = 1, 2, 3 \ldots s - 1 \quad (1)$$

Here control parameter can denote as μ . The value of μ is chosen as 4 as the particles move chaotically, β_i randomly generated number lies within range of 0, 1 and s presents population size . CMBA is the enhanced version of the Bat algorithm by including the features of chaotic variables in place of random variables used standard BA, which improvises searching operation as compared with stochastic search ergodicity, randomness, and dynamic motion properties of chaos. CMBA is the included version of the Chaos concept for population mapping and improvement of the diversity of immature swarm. The following equation expresses the mapping process.

$$x_i^c = \beta_i x_i \quad (2)$$

Here x_i^c presents ith bat location in the search space with chaotic perturbations and β_i represented as an ith chaotic sequence value.

3 Chaos-mapped bat algorithm (CMBA)

Xin-She Yang proposed a meta-heuristic algorithm derived from the nature of the echolocation nature of Bats. An echolocation is an approach adopted by the bats to search the prey. During the searching for the food, the bats emit short pulses, but when it found the prey is close the pulse emits rate increases and frequency becomes tuned up [18, 19]. Echolocation characteristics of bats characterized on the following three rules such as:

- Echolocation is the method adopted by the bats to sense the distance between the present position of the bat and the place of food located. The god-given magical way the bats calculate the distance even if barrier exists between two positions.
- Depending on the target of the food they emit the pulses rate within a range of 0, 1 and they fly with velocity (v_i) with position (x_i), frequency (f_{min}), and wavelength (λ).
- With a consideration that Loudness varies within a range of A_0 to A_{min}.

Chaotic mapping strategy is used in CMBA used to produce a high quality initial population with a enhanced quality which leads to fast convergence and outstanding final solution. Current position (x_i),velocity (v_i) of individual bat(i) is defined in d dimensional in the search space. The position of the bat changes with each iteration [14, 20, 21]. So at time instance t the position and velocity of the bat can be expressed as x_i^t and v_i^t respectively can be calculated using the following equations

$$f_i = f_{min} + (f_{max} - f_{min})\, Chaotic\, Mapping\, at\ ithiteration \quad (3)$$

$$v_i^t = v_i^{t+1} + (x_i^t - x^*)\, Chaotic\, Mapping\ \ ithiteration * f_i \quad (4)$$

$$x_i^t = x_i^{t-1} + v_i^t \quad (5)$$

In CMBA, β (of standard bat algorithm) represented as the chaotic iteration number varies between 0 and 1. x^* is the global best position during ith iteration among the total population of Bats. The parameter λ_i presented as chaotic iteration number within the range of 0, 1. Initially, each bat is randomly assigned a frequency that is drawn uniformly from f_{max},f_{min}. New solution x_{new} for each bat can be generated locally using:

$$x_{new} = x_{old} + \epsilon A^t \quad (6)$$

The value of ϵ chosen randomly within the range of $(-1,1)$.The Loudness (A) of the standard bat algorithm is placed with chaotic maps to enhance bat algorithm performance. Value of A changes in each iteration deferentially.To enhance the performance of BA pulse emission rate (r) of each bat substitute with chaotic maps.In comparison with standard BA, the pulse rate varies monotonically between 0 and 1, where as in EMBA, it replace with chaotic number between 0 and 1.

4 SVM and RBF Kernel

SVM is considered to be one of the best binary classification methods used by various researchers which works with a data sample with some predetermine kernel function and produce a class of the sample. The classification given problem is divided into training and test data sets, the generated output specifies the class label. An SVM maximizes the geometric margin between two classes of data and minimizes the error from false classified data points. The primal form of a soft-margin SVM can be written as

$$\min_{w \epsilon R^F} \frac{1}{2}[w]_2^2 + C\sum_{i=1}^{n} I(y_i, f_w(X_i)) \quad (7)$$

Here w the vector of the hyper plane used to differentiate two classes. $I(y, \hat{y})$ specifies the loss function in $\hat{y}$. C presents the regularization parameter for the evaluation of weight smoothness as well as errors. $f_w(x_i) = \langle \phi(x_i), w \rangle$ where $\phi(x) : R^d \rightarrow R^F$ presents the mapping function which maps the input data from the input space R^d to feature space R^F where F may be infinity. When the F

value becomes very large then the products in the inner space R^F can be computed using the kernel function $k(x, y) = \langle \phi x, \phi y \rangle$. Out of four basic SVM kernel functions (linear, polynomial, sigmoid, and radial basis functions (RBF)), in this study, we have preferred RBF as kernel function for SVM, Which is expressed as

$$k(x, y) = \frac{\exp ||x - y||^2}{2 \times \sigma^2} \tag{8}$$

To achieve the classification accuracy of the SVM better,different parameters such as C, V, and kernel parameters values should chosen carefully.

5 The proposed model: CMBA-SVM

This section describes the proposed Chaotic BAT algorithm SVM (CMBA-SVM) to uncover optimal values of SVM parameters. CMBA -SVM is trained with data sets such as Parkinson's data [22] set obtained from the UCI repository and Parkinson's disease Data sets were taken from Sakar et al. Istanbul [23–25]. Exploration and exploitation are two phases of the optimization algorithm solution. Exploration phases mean it explore search space to find out the optimal or near-optimal solutions and leads to new search space which may contain a better solution [26–28]. In the exploitation phase, the search locally near the current best solution explores the search space to find out a different solution. There should balance between random and greedy solutions in case of an optimization algorithm called exploitation while exploring the search space to find different solutions called exploration.

The proposed model, CMBA-SVM, it begins with the data prepossessing step. This data processing step is very important to (1) avoid features in greater numeric ranges dominating those in smaller ranges, (2) avoid during the computation numerical difficulties should be avoided, (3) achieve higher classification accuracy. For data processing feature is scaled to the range [0,1] as follows,

$$v' = \frac{Original\ value - lower\ bound\ value}{Upper\ bound\ value - Lower\ bound\ value} \tag{9}$$

Where:

- v is the original value from the given data set.
- Min and max present lower and upper bounds of the feature value, respectively.
- v' is the scaled value of the feature after normalization.

The whole normalized data is divided into two parts in (90% and 10%) ratio for training and test subsets. Then we have to define various parameters such as population no of bats, position or location of the bat in the search space, maximum iteration number during the execution, frequency, pulse rate, and velocity of each bat which is presented in the table. To train the SVM model we have used two parameters of Bat such as the value of C and the kernel parameter ς. From the two-parameter, it is clear that we have considered the search space of the given problem is a two dimensional. Each position of the bat swarm can be presented by the two-dimensional value of C and σ value. The position of each bat in the search space changes randomly and the position of the bat can be presented with parameters presented in Table 1. The search space increases due to the increment of the searching limits which helps to achieve the optimal solution. The main disadvantage is due to more no of computation the convergence rate decreases. For this study, we have considered the fitness function as stated below (Fig. 1).

$$Minimize\ function = \frac{Nos.\ of\ samples\ misclassified}{Total\ testing\ samples} \tag{10}$$

In CMBA, the positions of bats are modified according to Eqs. (3)–(6). The CMBA algorithm proceeds until it satisfied termination criteria, i.e. it reaches the maximum number of iterations or when the best solution is not updated for a given number of iterations.This structure utilizes the CMBA model to recognize the prospective combination of features from the data set by removing irrelevant and redundant information [28–31]. The CMBA identifies the best combination of features by adaptively searching the feature space and the learning accuracy of SVM determines the fitness value of the bats.

5.1 Algorithm 1. Pseudo-code of the chaotic mapped bat algorithm (CMBA)

1. Initialize the various parameters of bat such as positions(X_i), velocity(V_i), frequency(f_i), the pulse

Table 1 Initial parameters of the CMBA algorithm

Parameter details	Value
σ,α	0.9
$Loudness(A_0)$	0.9
r_0	0.5
f_{min}	0
f_{max}	1
c_{max}	35000
c_{min}	0.01
$sigma_{min}$	0.01
$sigma_{max}$	100
$Searchingranged_{min}$	100
d_{max}	30

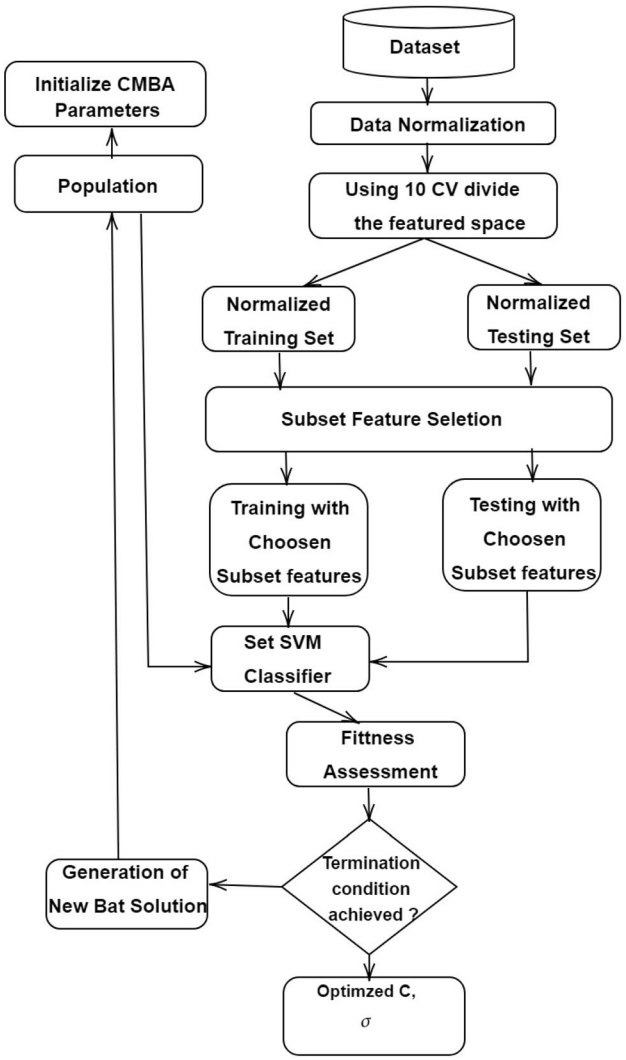

Fig. 1 Block diagram CMBA-SVM

emission rate(r_i), and loudness (A_i) for each bat. Consider the fitness function (F) is a minimum.
2. While (not meet up to stopping criteria) do
3. For individual bat process below steps
4. Using Eqs. (3)–(6) new solutions may be identify after updating (X_i), (V_i) and (f_i) value.
5. If($rand >$ emission rate)then
6. Local solution, X_{new} should be generate randomly, from the available best solution (exploitation) with in the search space.
7. End of if step 5
8. New solution (F_{new}) evaluated.
9. if ($rand <$ current bat loudness) and ($newbatlocation <$ F(X_i).
10. Current solution bat gets update.
11. Increase and decrease the value of (r_i) and (A_i) respectively.
12. End of step 9.
13. As per to their fitness function sorts the available bats and best solution (X_i) should be identified.
14. End of for loop step 3.
15. End of while loop.
16. Produce the best solution.

6 Experimental results and discussion

6.1 Description of experimental data

To obtain an unbiased comparison of CPU times, every experiment are performed using a single PC with hardware settings as shown in Table 2.

6.2 Data sets description

For the evaluation and effectiveness of the proposed CMBA-SVM model, we have considered two types of data sets such as (1) PD1 Parkinson's data was taken from the UCI machine learning repository (2) PD 2 Parkinson's disease Data sets were taken from Sakar et al. Istanbul.

6.2.1 UCI Parkinson's data

This data set is taken from the UCI Repository to classify from healthy disease patients from PD patients. These data sets contain details of biomedical voice measurement of 31 patients, out of this 23 no's are PD with eight healthy controls. Every instance of the data set contains a particular voice measurement of each patient. There are total voice recording of 195 patients with six different phonation of the vowel having a duration of 36 s. Samples of patients vary within the range of 0–28 years and 46–85 years with a mean of 65.8. From the data set description, it's clear that there is no missing value and contains real values presented in Fig. 3.

Table 2 Hardware and software requirement

Hardware	CPU name	Core (TM) i5-2400
	CPU frequency	3.10 GHz
	RAM	8 GB
	HDD	500 GB
Software	OS	Win 10
	Tool	MATLAB , WEKA
	Language	Java

6.2.2 Istanbul Parkinson's disease data

The data set during this research work was gathered from Sakar et al.from Istanbul, Turkey. This could be likewise called as Istanbul Parkinson's disease data set. It comprises of assorted sound recordings, incorporates continued vowels, numbers, words, and short sentences from 68 subjects. This training data set gathered from 40 persons including 20 no's every patient with PD within a spread of 43–77 and non PD patients within a spread of 45–83, though, testing data was comprised of 28 unique patients with PD within a variety 39 and 79. We favored just three sorts of sustained vowel recordings /a/, /o/, and /u/, practically like the Oxford PD data set. We merged them and framed a database that comprises of a total of 288 sustained vowel samples ',' and subsequently were prepared on these samples. Figure 3, a bunch of twenty-two and 26 linear and time-frequency based features are extracted for every voice sample of both forms of data sets.

6.3 Experimental parameter setting

For the implementation of our purposed model CMBA-SVM, we have used Java using WEKA API in Windows 10, Intel(R), Core (TM) i5-2400, and 8 GB RAM. For classification, we have used the LIBSVM package originally developed by Chang and Lin. The PD data set is normalized within the range of [0, 1] as the features with higher numerical ranges may dominate the smaller and it affects the classification performance. For the selection of featured genes from the PD data set, we have considered the wrapper model using RBF kernel-based SVM with ten cross-validations for evaluation of classification performance. The parameters of CMBA-SVM in mentioned in Table 1. Searching range of regularization and kernel parameter(C, σ) specified as $(2^{-5}, 2^{15})$ and $(2^{15}, 2^{5})$ respectively. Ten independent runs have taken for evaluation of the CMBA-SVM model and compared it with ECFA-SVM, CFA-SVM. Evaluation of the proposed method, done in terms of classification accuracy (ACC), sensitivity, specificity, and AUC.

6.4 Performance evaluation

For our experiment, we have used tenfold CV for evaluating the performance of proposed RBF-K-SVM classifier. The evaluation of the accuracy presents the selected featured genes form the adopted feature selection model. For each selection in the SVM model, we have used nested stratified tenfold CV for the construction of SVM classification model. Data set is divided within the ratio of (90–10%) whereas 90% of the data (9 fold-out of 10) taken for the training of the proposed model whereas 10% (1 fold-out of 10) of data used for testing the evaluation. The performance of the SVM model is presented with the help of receiver operating characteristic (ROC) curves. ROC curve presents a relationship graph about false positive rate and true positive rate of the error matrix. If true positive is 1 and false positive is 0 then curve occupies maximum area called area under the curve (ROC).ACC,AUC,sensitivity, and specificity were taken to evaluate the performance of different models. These measurements are defined as

$$ACC = \frac{True_{Pos} + True_{Neg}}{True_{Pos} + True_{Neg} + False_{Pos} + False_{Neg}} \tag{11}$$

$$Sensitivity = \frac{True_{Pos}}{True_{Pos} + False_{Neg}} \times 100 \tag{12}$$

$$Specificity = \frac{True_{Neg}}{True_{Pos} + False_{Neg}} \times 100 \tag{13}$$

where $True_{Pos}$ is the no's of true + ve , $False_{Neg}$ means the no's of −ve, $True_{Neg}$ represents the no's of true −ve, and $False_{Pos}$ is the no's of false + ve. AUC is the area under the ROC curve.

6.5 Results analysis

The proposed EMBA-SVM feature selection model is used to identify the featured genes from both PD data set was taken from the UCI machine learning repository and Sakar et al., Istanbul. The proposed EMBA-SVM framework was considered to implement experiments using a tenfold CV scheme to detect the most select subset of features. The performance comparison of the proposed method is done with CFA-SVM and ECFA-SVM. Figures 2 and 3 present the comparison graph accuracy of the proposed algorithm in the training phase of both data sets with 100 iterations and Loss during the training phase with 100 iterations respectively. From the literature survey with ECFA-SVM the same PD1 it achieves accuracy, sensitivity, precision, F-measure, specificity, area under ROC as 97.95%, 97.90%, 97.90%, 97.90%, 96.50%, 97.20% respectively whereas in GA-SVM (PD1) it achieves accuracy,sensitivity,Precision, F-measure, specificity, area under ROC as 87.69%, 87.70%, 88.40%, 87.90%, 83.40%, 85.50% respectively. But as mentioned in the comparison table , our proposed model performs better in both data sets with an accuracy of 98.24% (for PD1), and 99.49% (for PD2). The Error matrix of the proposed model is presented in Table 3. Classification accuracies comparison of other existing methods applied to the diagnosis of PD are presented Table 4 and Figs. 4, 5 and 6.

Oxford PD data set Label Feature			Istanbul PD Data Label Feature		
S1	MDVP: Fo(Hz)	Frequency Parameters	S1	Jitter(local)	Frequency Parameters
S2	MDVP: Fhi(Hz)		S2	Jitter(local, absolute)	
S3	MDVP: Flo(Hz)		S3	Jitter(RAP)	
S4	MDVP: Jitter(%)		S4	Jitter(PPQ 5)	
S5	MDVP: Jitter(Abs)		S5	Jitter(DDP)	
S6	MDVP: RAP		S6	Number of pulses	Pulse Parameters
S7	MDVP: PPQ		S7	Number of periods	
S8	Jitter: DDP		S8	Mean period	
S9	MDVP: Shimmer	Amplitude Parameters	S9	Standard dev. of period	
S10	MDVP:Shimmer(dB)		S10	Shimmer(local)	Amplitude Parameters
S11	Shimmer: APQ3		S11	Shimmer(local, dB)	
S12	Shimmer: APQ5		S12	Shimmer(APQ3)	
S13	MDVP: APQ		S13	Shimmer(APQ5)	
S14	Shimmer: DDA		S14	Shimmer(APQ 11)	
S15	NHR	Voicing Parameters	S15	Shimmer(DDA)	
S16	HNR		S16	Fraction of unvoiced frames	Voicing Parameters
S17	RPDE		S17	Number of voice breaks	
S18	D2	Fractal scaling	S18	Degree of voice breaks	
S19	DFA		S19	Median pitch	Pitch Parameters
S20	Spread1	Pitch Parameters	S20	Mean pitch	
S21	Spread2		S21	Standard deviation	
S22	PPE		S22	Minimum pitch	
			S23	Maximum pitch	
			S24	Autocorrelation	Harmonocity Parameters
			S25	Noise-to-Harmonic	
			S26	Harmonic-to-Noise	

Fig. 2 Data set details

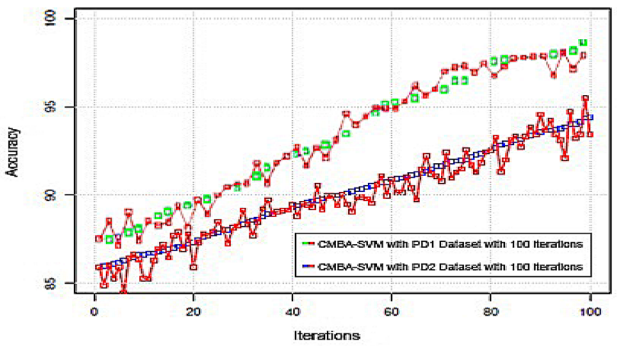

Fig. 3 Accuracy for CMBA-SVM training phase with 100 iterations

7 Conclusion

We have projected a hybrid instance-based approach based on a chaotic BA and applied it to identify the PD affected persons from a healthy one from two sets of skewed PD data sets. RBF-SVM was included with CMBA for the correct prediction of the desired output during the execution of the program. The result The results suggested that the proposed CMBA-SVM model performed outstandingly in comparison with a various met heuristic approach like CBA-SVM and ECBA-SVM, in terms of various performance metrics. We hope the proposed model performs

Table 3 Error matrix of the proposed model

Model	Accur-acy	Sensi-tivity	Precision	F-Measure	Specif-icit	ROC area
$ECFA - SVM(PD1)$	97.95	97.90	97.90	97.90	96.5	97.20
$GA - SVM(PD1)$	87.69	87.70	88.40	87.90	83.40	85.5
$CMBA - SVM(PD1)$	98.24	98.1	98.56	98.08	97.78	98.79
$CMBA - SVM(PD2)$	99.49	98.11	98.83	98.91	98.15	99.11

Table 4 Comparsion of proposed model accuracy with existing one

Study	Method	Accuracy
[32]	CBFO-SVM	97.89
[33]	CFS-RF	87.10
[34]	SVM-BFO	97.42
[35]	PCA-FKNN	96.07
With PD1	CMBA-SVM	98.24
With PD2	CMBA-SVM	99.49

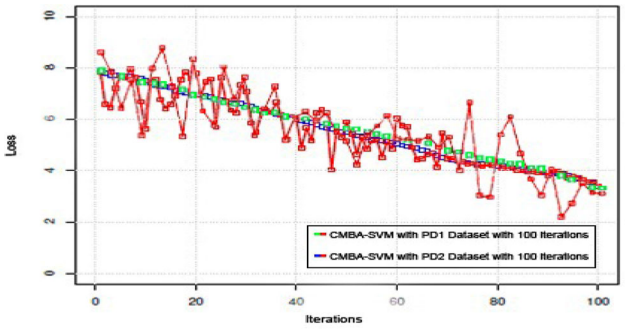

Fig. 4 Loss for CMBA-SVM training phase with 100 iterations

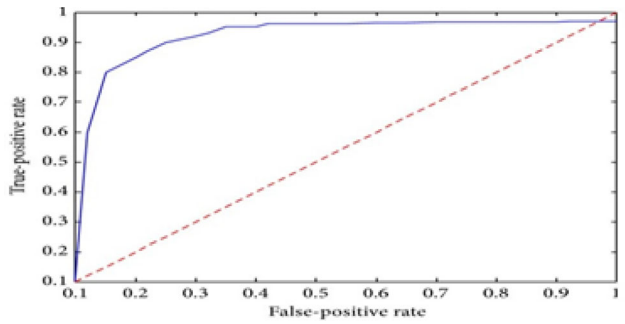

Fig. 5 ROC graph for PD1 data set using CMBA-SVM

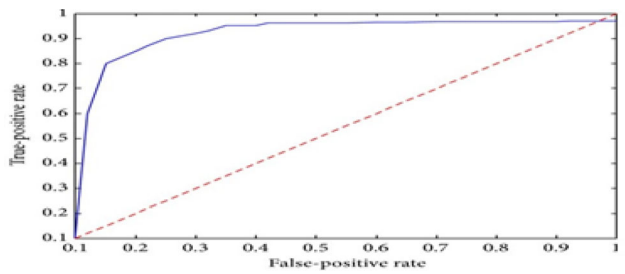

Fig. 6 ROC graph for PD2 data set using CMBA-SVM

better up to a good extent. Finally, we will generalize the proposed method to much larger data sets in the future.

References

1. Sahu B, Panigrahi A, Sukla S, Biswal BB (2020) MRMR-BAT-HS: a clinical decision support system for cancer diagnosis. Leukemia 7129(73):48
2. Sahu B, Panigrahi A, Mohanty S, Sobhan S (2020) A hybrid cancer classification based on SVM optimized by PSO and reverse firefly algorithm. Int J Control Autom 13(4):506–517
3. Sahu B, Mohanty S, Rout S (2019) A hybrid approach for breast cancer classification and diagnosis. EAI Endorsed Trans Scalable Inf Syst 6(20)
4. Sahu B (2019) Multi-tier hybrid feature selection by combining filter and wrapper for subset feature selection in cancer classification. Indian J Sci Technol 12:3
5. Sahu B, Dash S, Nandan Mohanty S, Kumar Rout S (2018) Ensemble comparative study for diagnosis of breast cancer datasets. Int J Eng Technol 7(4.15):281–285
6. Sahu B (2018) A combo feature selection method (filter+ wrapper) for microarray gene classification. Int J Pure Appl Math 118(16):389–401
7. Sahu B, Panigrahi A, Pani S, Swagatika S, Singh D, Kumar S (2020) A crow particle swarm optimization algorithm with deep neural network (CPSO-DNN) for high dimensional data analysis. In: 2020 International conference on communication and signal processing (ICCSP), IEEE, pp 0357–0362
8. Prashanth R, Roy SD, Mandal PK, Ghosh S (2014) Automatic classification and prediction models for early Parkinson's disease diagnosis from SPECT imaging. Expert Syst Appl 41(7):3333–3342
9. Sarwar A, Ali M, Manhas J, Sharma V (2020) Diagnosis of diabetes type-II using hybrid machine learning based ensemble model. Int J Inf Technol 12(2):419–428
10. Nandhini N, Thangadurai K (2019) An incremental rough set approach for faster attribute reduction. Int J Inf Technol 36:1–15
11. Prasanth A, Valsala S (2017) Semantic chameleon clustering analysis algorithm with recommendation rules for efficient web usage mining. In: 2017 9th IEEE-GCC conference and exhibition (GCCCE), IEEE, pp 1–9
12. Sahu B, Badajena JC, Panigrahi A, Rout C, Sethi S (2020) 7 An intelligence-based health biomarker identification system using microarray analysis. Applied intelligent decision making in machine learning. CRC Press, Cambridge, pp 137–161
13. Chen M, Hao Y, Hwang K, Wang L, Wang L (2017) Disease prediction by machine learning over big data from healthcare communities. IEEE Access 26(5):8869–79
14. John Prince, Fernando Andreotti, Maarten De Vos (2018) Multi-source ensemble learning for the remote prediction of Parkinson's disease in the presence of source-wise missing data. IEEE Trans Biomed Eng 66(5):1402–11
15. Darwish A (2018) Bio-inspired computing: agorithms review, deep analysis, and the scope of applications. Future Comput Inform J 3(2):231–46
16. Lahmiri S, Shmuel A (2019) Detection of Parkinson's disease based on voice patterns ranking and optimized support vector machine. Biomed Signal Process Control 49:427–33
17. Sahu B (2019) Multi filter ensemble method for cancer prognosis and Diagnosis. Int J Eng Appl Sci Technol 4:105–109
18. Polat K (2019) A hybrid approach to Parkinson disease classification using speech signal: the combination of smote and random forests. In: 2019 Scientific meeting on electrical-electronics and biomedical engineering and computer science (EBBT), IEEE, pp 1–3.
19. Gupta D, Sundaram S, Khanna A, Hassanien AE, De Albuquerque VHC (2018) Improved diagnosis of Parkinson's disease

using optimized crow search algorithm. Computers & Electrical Engineering, 68, 412–424
20. Haijun L, Zhongwei H, Feng Z, Ahmed E, Ee-Leng T, Hancong L, Jing Q, Baiying L (2018) Parkinson's disease diagnosis via joint learning from multiple modalities and relations. IEEE J Biomed Health Inform 23(4):1437–49
21. Nilashi M, Ibrahim O, Samad S, Ahmadi H, Shahmoradi L, Akbari E (2019) An analytical method for measuring the Parkinson's disease progression: a case on a Parkinson's telemonitoring dataset. Measurement 136:545–57
22. Haq AU, Li JP, Memon MH, Khan J, Malik A, Ahmad T, Ali A, Nazir S, Ahad I, Shahid M (2019) Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson's disease using voice recordings. IEEE Access 7:37718–37734.
23. Sakar BE, Isenkul ME, Sakar CO, Sertbas A, Gurgen F, Delil S, Apaydin H, Kursun O (2013) Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE J Biomed Health Inform 17(4):828–34
24. Ghosh S. Identifying click baits using various machine learning and deep learning techniques
25. Mittal K, Aggarwal G, Mahajan P (2019) Performance study of K-nearest neighbor classifier and K-means clustering for predicting the diagnostic accuracy. Int J Inf Technol 11(3):535–540
26. Mohammad RS, Mojtaba S, Abdollah S, Saeed S, Ivan SK, Vesna S, Arman R (2019) Optimized machine learning methods for prediction of cognitive outcome in Parkinson's disease. Comput Biol Med 111:103347
27. Salmanpour MR, Shamsaei M, Saberi A, Setayeshi S, Taherinezhad E, Klyuzhin IS, Tang J, Sossi V, Rahmim A (2018) Machine learning methods for optimal prediction of outcome in Parkinson's disease. In: IEEE nuclear science symposium and medical imaging conference proceedings (NSS/ MIC), IEEE, pp 1–5
28. Shahab S, Somayeh H, Vahdat S, Asl IM, Kazemipoor M, Rabczuk T (2019) Parkinson's disease detection using biogeography-based optimization. Comput Mater Contin 61:11–26
29. Xu Y, Chen H, Heidari AA, Luo J, Zhang Q, Zhao X, Li C (2019) An efficient chaotic mutative moth-flame-inspired optimizer for global optimization tasks. Expert Syst Appl 129:135–55
30. Tharwat A, Hassanien AE, Elnaghi BE (2017) A BA-based algorithm for parameter optimization of support vector machine. Pattern Recognit Letters, 93:13–22
31. Juneja K, Rana C (2018) An improved weighted decision tree approach for breast cancer prediction. Int J Inf Technol 12:1–8
32. Cai Z, Gu J, Wen C, Zhao D, Huang C, Huang H, Chen H (2018) An intelligent Parkinson's disease diagnostic system based on a chaotic bacterial foraging optimization enhanced fuzzy KNN approach. Comput Math Methods Med 2018:24
33. Ozcift A, Gulten A (2011) Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. Comput Methods Progr Biomed 104(3):443–451
34. Cai Z, Gu J, Chen HL (2017) A new hybrid intelligent framework for predicting Parkinson's disease. IEEE Access 5:17188–17200
35. Chen HL, Huang CC, Yu XG, Xu X, Sun X, Wang G, Wang SJ (2013) An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. Expert Syst Appl 40(1):263–271